

ABSTRAK

Distribusi zakat bersifat sekuensial dan menghadapi tantangan dalam mengoptimalkan kesejahteraan mustahik, menjaga keadilan jangka panjang, serta mematuhi aturan institusional secara simultan, sementara pendekatan yang ada belum mengintegrasikan ketiga aspek tersebut. Penelitian ini bertujuan merancang dan mengevaluasi pendekatan Hybrid Long-Term Fairness (HLTF) sebagai Constrained Markov Decision Process (CMDP) untuk kebijakan distribusi zakat. HLTF diselesaikan menggunakan algoritma PPO-Lagrangian yang mengintegrasikan Long-Term Benefit Rate (LBR) sebagai cost constraint eksplisit dan action projection untuk menjamin kepatuhan aturan institusional, dievaluasi pada lingkungan simulasi semi-sintetis berbasis data BAZNAS dan BPS Kota Bandung. Hasil menunjukkan HLTF mencapai reward rata-rata 0,086665 dengan violation rate 1,64%, turun 98,22% dibandingkan PPO Vanilla (91,80%). Pada evaluasi 250 episode *test-split held-out*, HLTF mempertahankan performa serupa dengan reward 0,086748 dan violation rate 6,95%, sementara pada skenario 2 (*shock*) yang belum pernah dilihat, generalisasi menurun violation rate 76,92%, menunjukkan arah pengembangan lanjutan.

Kata Kunci: Distribusi Zakat, *Reinforcement Learning*, *Constrained Markov Decision Process*, *Proximal Policy Optimization*, *Long-Term Fairness*, *Hybrid Long-Term Fairness*.